

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



14 FEB 2005



(43) International Publication Date
4 March 2004 (04.03.2004)

PCT

(10) International Publication Number
WO 2004/019264 A1

(51) International Patent Classification⁷:
G06F 19/00 // 159:00

G06K 9/62,

(74) Agent: ELLA CHEONG MIRANDAH & SPRUSONS
PTE LTD; Robinson Road Post Office, P.O. Box 1531,
Singapore 903031 (SG).

(21) International Application Number:

PCT/SG2002/000190

(22) International Filing Date: 22 August 2002 (22.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): LAB-
ORATORIES FOR INFORMATION TECHNOLOGY
[SG/SG]; 21 Heng Mui Keng Terrace, Singapore 119613
(SG).

(72) Inventor; and

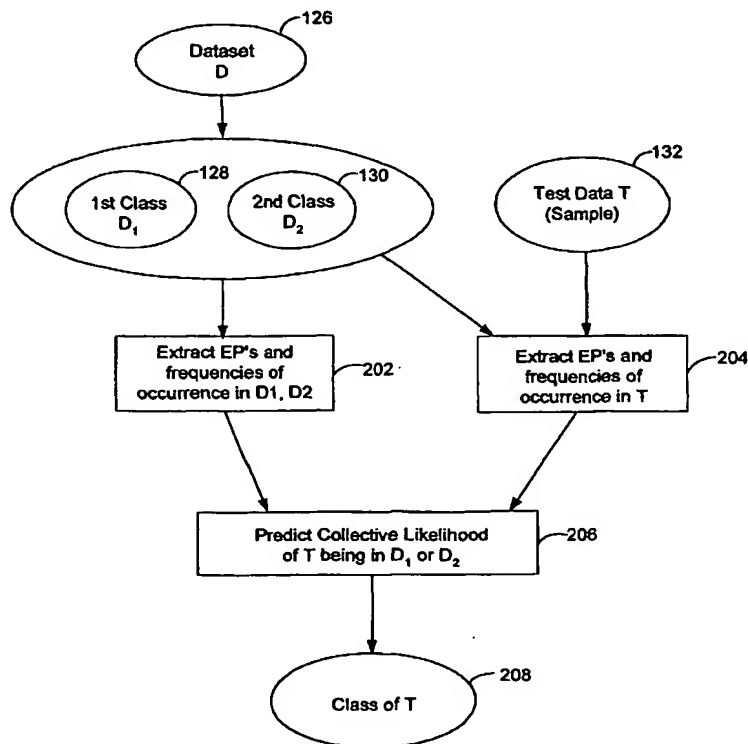
(75) Inventor/Applicant (for US only): LI, Jin, Yan [AU/SG];
19B, Hillview Avenue, #11-04, Hillview Park, Singapore
669555 (SG).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK,
TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: PREDICTION BY COLLECTIVE LIKELIHOOD FROM EMERGING PATTERNS



(57) Abstract: A system, method and computer program product for determining whether a test sample is in a first or a second class of data (for example: cancerous or normal), comprising: extracting a plurality of emerging patterns from a training data set, creating a first and a second list containing respectively, a frequency of occurrence of each emerging pattern that has a non-zero occurrence in the first and in the second class of data; using a fixed number of emerging patterns, calculating a first and second score derived respectively from the frequencies of emerging patterns in the first list that also occur in the test data, and from the frequencies of emerging patterns in the second list that also occur in the test data; and deducing whether the test sample is categorized in the first or the second class of data by selecting the higher of the first and the second score.

WO 2004/019264 A1